Insilico Analysis & 3D Structure Prediction of Hemagglutinin Protein of H9N2 Avian Influenza Virus

Samraddhi Tiwari
Department of Biotechnology
Madhav Institute of Technology & Science
Gwalior, India

Vinod Kumar Jatav
Asst. Professor, Department of Biotechnology
Madhav Institute of Technology & Science
Gwalior, India

***Abstract*— H9N2 avian influenza virus are widespread in chickens,quail and other poultry in asia and have caused a few cases of influenza in humans. To understand the structural and functional analysis of H9N2 avian influenza viruses this study was done. We studied 29 amino acid sequences of haemagglutinin (HA) genes, which were gained from different 8 countries from the year 2004 to 2012. HA is a surface protein which is responsible for the mode of infection. Phylogenetic tree showed the differentiation and similarity between the HA amino acid sequences of strains. On the basis of phylogenetic result two sequences ACF93484 (Gujrat) and AEQ33497 (Karachi) were selected for the further analysis. Physiochemical characterization was done to interpret properties like pI, EC, AI, GRAVY and instability indexing. The 3D structure of this protein is not available so homology modeling was performed to generate good quality models. For the verification of protien procheck was used. The predicted model can be used in structure based drug designing and vaccine development.**

Keywords- Avian Influenza Virus (AIV); Haemagglutinin (HA); Grand Average hydropathy (GRAVY); Insilico Analysis; 3D Structure Prediction.

I. INTRODUCTION

H9N2 is a subtype of the species Influenza A virus (bird flu virus) (Murphy et al., 1996). H9N2 influenza virus are widespread in chickens,quail and other poultry in asia and have caused a few cases of influenza in humans (Mikhail N .Matrosovich). In April 1999, two World Health Organization reference laboratories independently confirmed the isolation of avian influenza A (H9N2) viruses for the first time in humans (Timothy et al., 1999). Avian influenza A viruses (AIV) are enveloped, segmented and negative-stranded RNA viruses. The subtypes of influenza A virus HA and NA (H1-H16 and N1-N9) is circulating in water birds, especially in migratory ducks (Azeem et al., 2010). Hemagglutinin (HA) is one kind of important AIV glycoprotein on the surface of H9N2 AIV. HA plays a key role in the process of virus absorption and trans-membrane control, and the neutralizing antibody produced by the simulated body can neutralize the infection of H9N2 subtype AIV (Yin Dai).HA is involved in the early stages of infection ,causing the binding of the sialic acid receptor present on the host cell surface and leading to fusion of the viral and endosomal membrane and subsequent entry into the host cell (Webster et al., 1992). Although avian influenza A viruses usually do not infect humans, rare cases of human infection with avian influenza A viruses have been reported. Evidence for five additional human illnesses attributed to H9N2 in Guangdong Province, China, during 1998 has been reported (Guo et al., 1999). Detection of antibody to H9N2 has been reported from persons in northern and southern China and poultry workers in Hong Kong (Eick et al., 2000), sug- gesting that additional unrecognized human H9N2 infections have occurred. Most human infections with avian influenza. Viruses have occurred direct contact with infected poultry. The signs and symptoms of avian influenza in humans have ranged from eye infections (conjunctivitis) to influenza-like illness symptoms (e.g., fever, cough, sore throat, muscle aches) to severe respiratory illness (e.g. pneumonia, acute respiratory distress, viral pneumonia) sometimes accompanied by nausea, diarrhea, vomiting and neurologic changes. CDC and WHO recommend oseltamivir, a prescription antiviral medication, for treatment and prevention of human infection with avian influenza A viruses. H9N2 influenza A virus from poultry in Asia have human virus like receptor specifity.

II. MATERIALS & METHODS

A. Retrieval of target sequence:

From the NCBI database, 29 amino acid sequences of the HA gene of influenza A virus used to study were retrieved from different countries during 2004-2012, corresponding to accession number ACP50620.1, ACP50741.1, AEQ33497.1, ACP50719.1, AFI73234.1, AFV68520.1, AFD62263.1, AFI73231.1, AFO82965.1, ACF93484.1, ADI79229.1, ACX55913.1, ADC30121.1, AEA76366.1, AEA76395.1, AFH53541.1, AFO83272.1, AFO83273.1, AAT37508.1, AAS48383.1, ABP48862.1, AAS48381.1, ABP48873.1, AAS48382.1, CBI68714.1, ACY25803.2, AFO83277.1, ABP48875.1 and ABP88149.1.

B. Phylogenetic analysis:

Phylogenetic patterns of Ha of H9N2 influenza virus were aligned by CLUSTAL W. It calculates the best matches for the

selected sequences, and lines them up so that the identities, similarities and differences can be seen. Polygenetic tree of 29 amino acid sequence were shown in fig. 1("http://www.ebi.ac.uk/Tools/msa/clustalw2").

C. Physico-chemical characterization:

The physical and chemical properties of Indian and Pakistan region were checked by PROT- PARAM. The values of theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient (Gill et al., 1989), instability index (Guruprasad et al., 1990) aliphatic index (Ikai et al., 1980) and grand average hydropathy (GRAVY) were computed. The parameters are shown in table 1. (http://web.expasy.org/protparam).

D. Secondary structure:

Secondary structure has been predicted using PSIPRED software where the FASTA format of the sequence was used. Secondary structure of Indian and Pakistan sequence were shown in fig 3. SOPMA (Geourjon et al., 1995) was employed for calculating the secondary structural features of the protein sequence considered for this study. The results are presented in Table 2.

E. 3D structure and Quality assessment:

Geno 3d is an automatic web server for protein molecular modelling (Combet et al., 2002) (http://geno3d-pbil.ibcp.fr).The stereochemical property of the protein was assessed by Ramchandran plot analysis using PROCHECK (http://nihserver.mbi.ucla.edu/SAVES/).The result is shown in Table no 3.

III. RESULTS & DISCUSSION

In this study, the protein sequence of Haemagglutinin protein of H9N2 influenza virus was retrieved from NCBI Entrez sequence search in FASTA format .Phylogenetic tree constructed of all 29 sequences of Haemagglutinin protein of H9N2 influenza virus, the rooted structure show the homologous sequences, orthologous sequences and paralogous sequence of influenza virus isolated from different countries and from the tree ,two sequences ACF93484.1 and AEQ33497.1 were selcted on the basis of their percentage of similarity.

Physiochemical Parameters computed using Expasy's ProtParam tool is represented in a table -1. If a protein is having instability index smaller than 40 than it is predicted as stable, on the other hand a value above 40 predicts that the protein may be unstable (Guruprasad et al., 1990) Instability index of both the sequences are 36.54 and 35.38, it indicates the stable nature of protein. The aliphatic index is considered as a positive factor for the increase of thermal stability. High aliphatic index (86.34 and 86.95) of query protein suggests that the protein may be stable for a wide temperature range. The Grand Average hydropathy (GRAVY) value is low (-0.324 and -0.329) it indicates the possibility of better interaction with water.

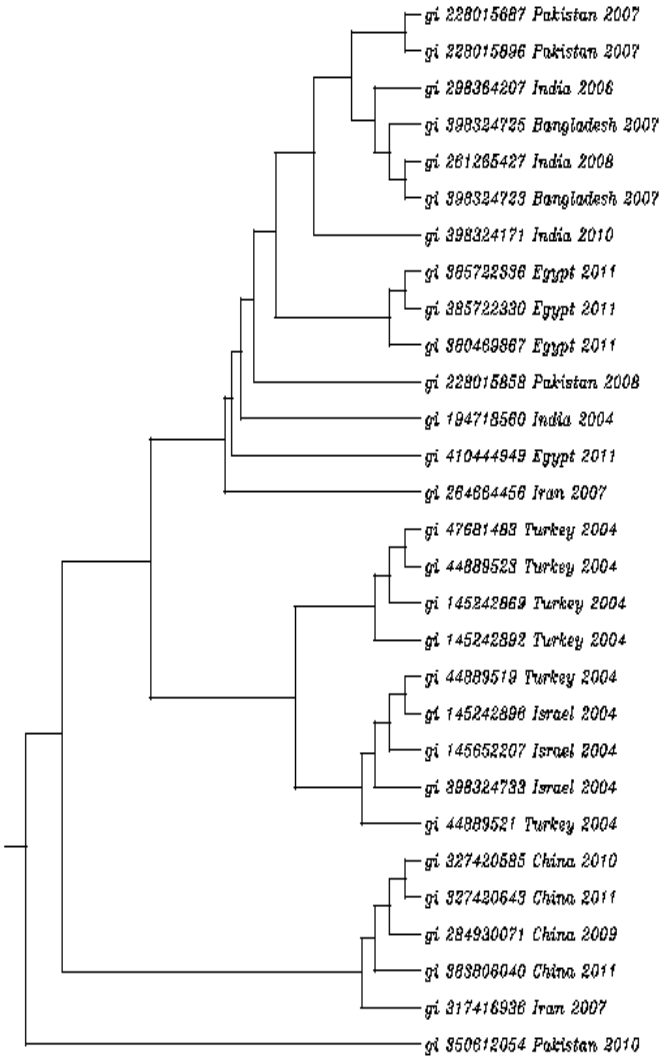

Figure 1. Phylogenetic analysis of 29 sequences of HA protein.

TABLE I. PARAMETERS COMPUTED USING EXPASY'S

Properties	Indian	Pakistan
No of amino acids	560	564
Molecular weight	62892.3	63253.6
Theoretical P.I	7.55	6.10
Total no of –ve (Asp + Glu)	54	62
Total no of +ve(Arg + Lys)	55	56
Extinction coefficients	90020	86830
Extinction coefficients*	89270	85830
Instability index	36.54	35.38
Aliphatic index	86.34	86.95
Grand average of hydropathicity	-0.324	-0.329

TABLE II. CALCULATED SECONDARY STRUCTURE ELEMENTS BY SOPMA

S.no	Parameters	Indian value(%)	Pakistan value(%)
1	Alpha helix	32.50%	32.80%
2	3₁₀ helix	0.00%	0.00%
3	Pi helix	0.00%	0.00%
4	Beta bridge	0.00%	0.00%
5	Extended strand	22.32%	20.21%
6	Beta turn	6.07%	7.27%
7	Bend region	0.00%	0.00%
8	Random coil	39.11%	39.72%
9	Ambiguous state	0.00%	0.00%
10	Other state	0.00%	0.00%

The secondary structure of Avian influenza virus of haemagglutinin protein was predicted by two software namely SOPMA (Self Optimized Prediction Method with Alignment) and PSI PRED. SOPMA predicts 69.5% of amino acids correctly to describe secondary structure prediction **(Geourjon *et al.*, 1995)**.The results of SOPMA are presented in Table-2. These results show higher number of random coils in comparison to other secondary structure elements (alpha helix, extended strand and beta turns), default parameters (Window width: 17, similarity threshold: 8 and number of states: 4) were taken by SOPMA for secondary structure prediction. Secondary structure and disorder prediction was made using PSI-PRED which is shown in figure 2 and fig 3.

Three dimensional structures of proteins are predicted due to unavailability of such data. There is no experimental structure found for the protein considered. The homology modelling of the protein was done by Geno3D. The results obtained from this program were compared in table 3. Finally model was visualized by Rasmol (figure 4 and 5).

The evaluation of predicted structure generated by Geno 3D for the stereochemical quality was done using Ramachandran map calculations done with the PROCHECK (figure 5 and 6). The 42.1% of Indian region and 54.6% of Pakistan region of residues were found in the core right handed alpha helices(A),beta sheets (B)and left-handed alpha helix(L) region. 39.4% of Indian and35.7% of Pakistan region residues were found in the allowed right- handed alpha helix(a),beta sheets(b) and left -handed alpha helices regions. The 10.5% and 6.7% of the residues were found in the generously allowed alpha helices (~a), beta sheets (~b), left handed alpha helices (~l) and epsilon (~p) regions. The 8% and 3% of the residues was found to be localized at the disallowed regions. The results indicate to a good quality of predicted model.

TABLE III. RAMACHANDRAN PLOT CALCULATION COMPUTED WITH THE PROCHECK PROGRAM

S.No.	Parameters	Indian value (%)	Pakistan Value (%)
1	Residues in the most Favoured Region	42.1	54.6
2	Residues in additionally allowed region	39.4	35.7
3	Residues in generously allowed region	10.5	6.7
4	Residues in disallowed region	8	3

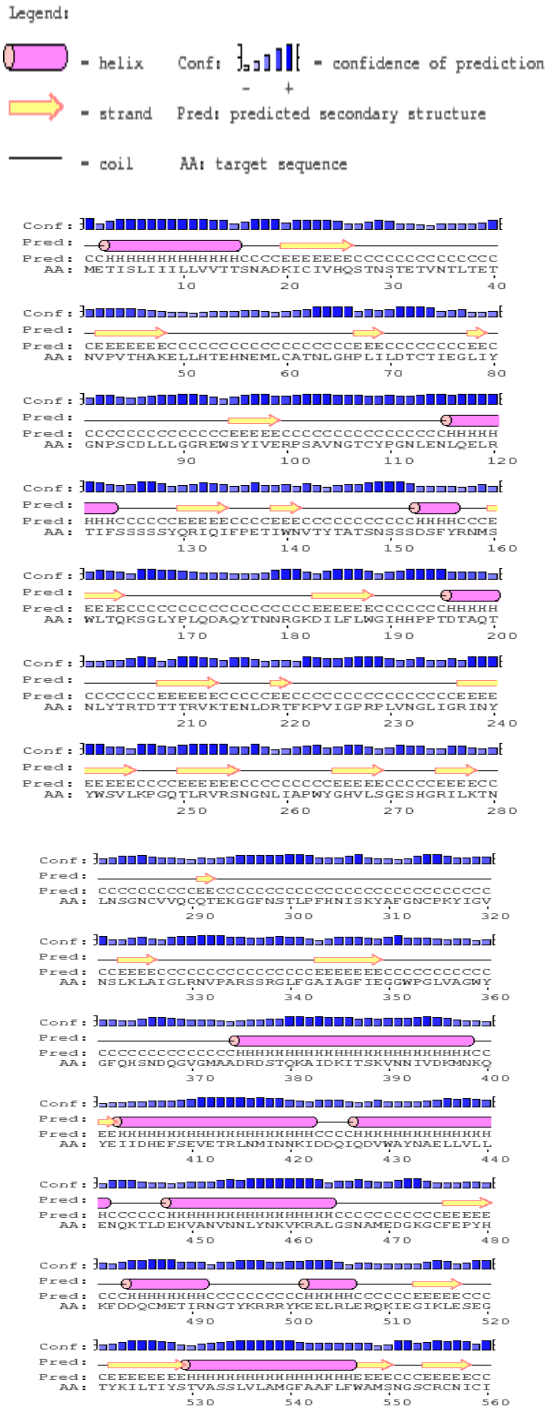

Figure 2. Indian Sequence (Sequence length is 560)

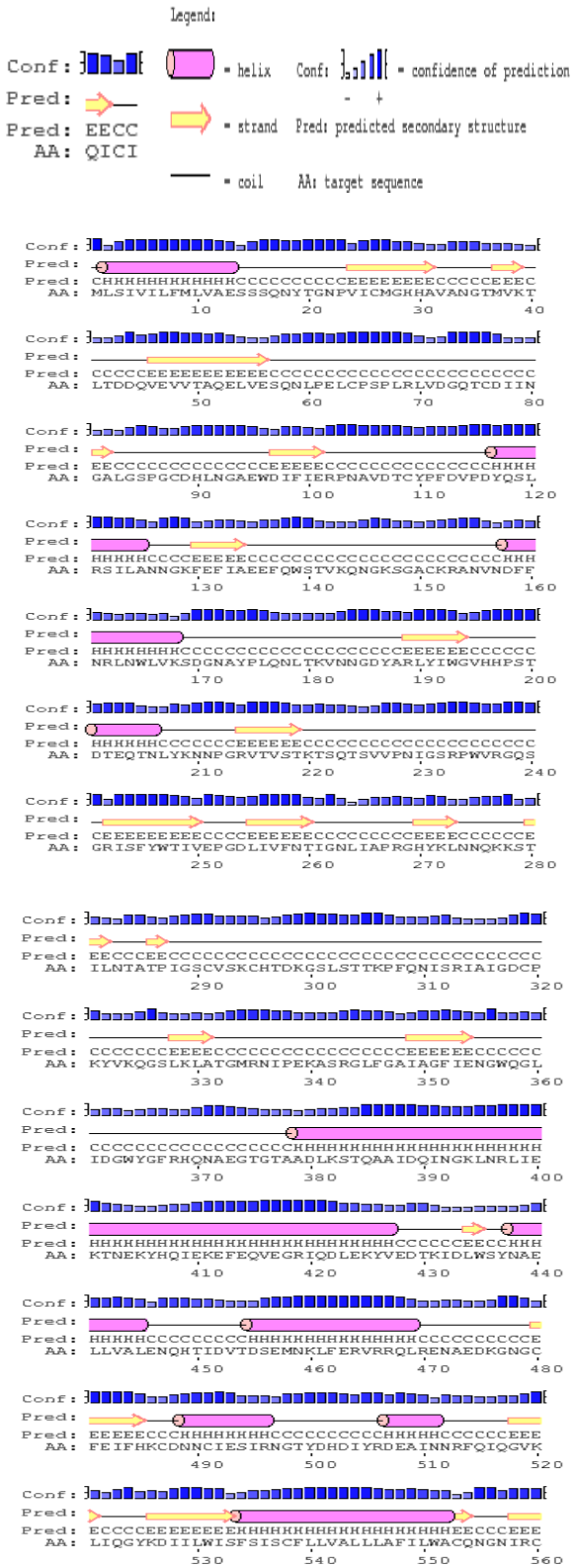

Figure 3. Pakistan Sequence (Sequence length 564)

Secondary structure of selected sequence was formed by the PSI PRED ,in this secondary structure cyndrical pink colour shows the helix ,the arrows shows the strand and the line shows the coil.

A. *TERTIARY STRUCTURE*

1) *Indian sequence*

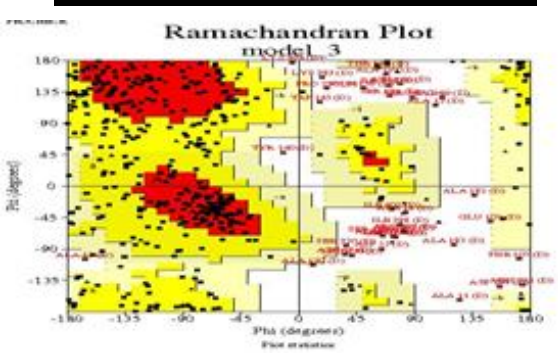

Figure 4. 3D structure and Ramachandran plot of Indian sequence formed by Geno 3D.

2) *Pakistan sequence*

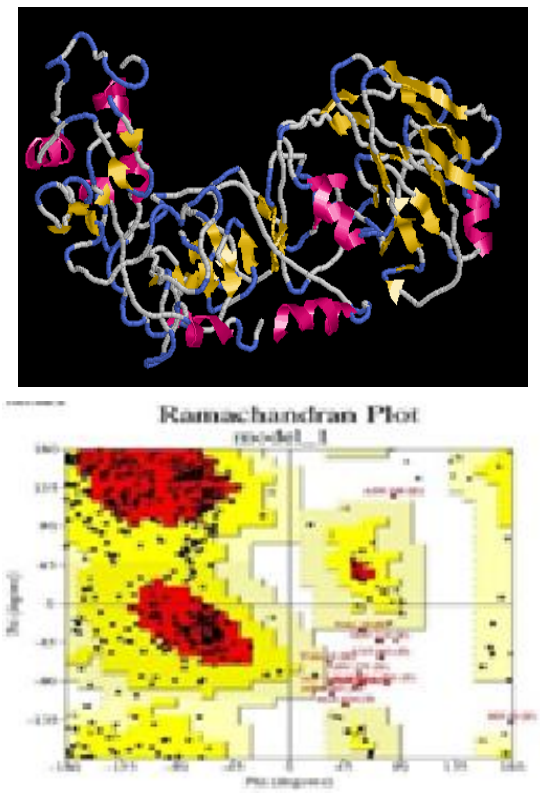

Figure 5. 3D structure and Ramachandran plot of Pakistan sequence formed by Geno 3D.

IV. CONCLUSION

On the basis of various structural and physiochemical parameters assessment, it can be concluded that the predicted three dimensional structure of Haemagglutinin protein of Avian influenza virus of both the countries are stable. Structural information of this model can be effectively used and can be further implemented in future drug designing.

ACKNOWLEDGEMENT

We would like to express our heartfelt feeling of gratefulness to all those who directly or indirectly helped us in this work.

REFERENCES

[1] Antoinette J. Piaggio, Susan A. Shriner, Kaci K. VanDalen, Alan B. Franklin, Theodore D. Anderson, Sergios-Orestis Kolokotronis,Molecular Surveillance of Low Pathogenic Avian Influenza Viruses in Wild Birds across the United States: Inferences from the Hemagglutinin Gene, World Health Organization. Influenza Weekly Epidemiological Record (1997), 74:111.

[2] Butt A.M., Siddique S., Muhammad I., Tong, Y. Avian influenza A(H9N2):computational molecular analysis and phylogenetic characterization of viral surface proteins isolated 1997 and 2009 from the human population, Virology jour. (2010).7:319.

[3] Combet C, Jambon M, Deleage G, Geourjon C ,Geno3D: Automatic comparative molecular modelling of protein. Bioinformatics ,(2002) ,18: 213-214.

[4] Dai Y,Wang J,Xu F, Zhao R, Hu X and Zhang D Sequence analysis of the haemagglutinin genes of H9N2 avian influenza viruses isolated in China during 1996 to 2011 , Food, Agriculture and Environment (JFAE), (2012),10(1):143-146.

[5] Eick A, Hu-Primmer J, Rowe T, Masseoud F, Fukuda K, Lim W, et al ,Seroprevalence of antibody to H9N2 viruses in poultry workers of Hong Kong. Poster #52, Session 7, July 16, 2000. International Conference on Emerging Infectious Diseases 2000, July 16-19, 2000, Atlanta, GA.

[6] Geourjon C, Deléage G (1995) SOPMA: signifi cant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 11: 681-684.

[7] Gill SC, Von Hippel PH ,Extinction coeffi cient. Anal Biochem ,(1989)182: 319-328.

[8] Guruprasad K, Reddy BVP, Pandit MW ,Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Prot Eng,(1990) 4: 155-164.

[9] Guo YJ, Li J, Cheng X, Wang M, Zhou Y, Zhou Y, et al., Discovery of man infected by avian influenza virus, Chin J Exp Clin Virol,(1999)13:105-8.

[10] Ikai AJ,Thermo stability and aliphatic index of globular proteins. J Biochem (1980)88: 1895- 1898.

[11] Murphy, B.R., Knipe, D.M. and Webster, R.G. Orthomyxoviruses. In Field's Virology, (1996),1397-1445.

[12] Shortridge KF,Pandemic influenza: A zoonosis? Semin Respir Infect ,(1992);7:11-25.

[13] Uyeki M.T, Chong Y, Katz.M.J, Lim W, Ho Y.Y, Wang S.S, Tsang T, Winnie W Y A, et al, Lack of Evidence for Human-to- Human Transmission of Avian Influenza A (H9N2) Viruses in Hong Kong, (1999)China.

[14] Webster R. G.,Bean, W. J., Gorman, O. T., Chambers, T. M., Kawaoka, Y.Evolution and ecology of influenza A viruses .Microbiol Rev ,(1992),56,152-179.